* Optimization Tools : 1) Lagrange's Multiplier
2) Newton - Raphson Method
3) Gradient descent Method

General Case : Maximize / Minimize f subject to constraint $g(x_1, \dots, x_n) = 0$ by solving the following equations simultaneously
$\nabla \{ f + \sum_{i=1}^{k} \lambda_i g_i \} = 0$

Example : $f(x, y) = x^2 y$; $g(x, y) = x^2 + xy = 12$

$F(x, y) = x^2 y + \lambda (x^2 + xy - 12)$

$\frac{\partial F}{\partial x} = 2xy + \lambda (2x + y) = 0$ (L1)

$\frac{\partial F}{\partial y} = x^2 + \lambda x = 0$ (L2)

(L1) : $x (x + \lambda) = 0 \Rightarrow x = -\lambda$ as $x = 0$ does not verify g

(L2) : $-2\lambda y + \lambda(-2\lambda + y) =$
$-2\lambda y - 2\lambda^2 + \lambda y =$
$-2\lambda^2 - \lambda y =$
$\lambda \neq 0$ $\quad 2\lambda + y = 0$

$\Rightarrow y = -2\lambda$

$g(-\lambda, -2\lambda) = \lambda^2 + 2\lambda^2 - 12$
$= 3\lambda^2 - 12$

$$= 3\lambda^2 - 12$$
$$= 3(\lambda^2 - 4)$$
$$= 3(\lambda - 2)(\lambda + 2) = 0$$

$\Rightarrow \lambda = 2$ or $\lambda = -2$

Hence, $(2, 4)$ and $(-2, -4)$ are the optimal points

$f(2,4) = 4(4) = 16$ maximum
$f(-2,-4) = 4(-4) = -16$ minimum

* Convergence Theory :

Let $X_1, \dots, X_n$ be a sequence of iid RVs

1) Independent and identically distributed :
$\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$

2) Strong Convergence :
(X_n) converges to X a.s $\Leftrightarrow \mathbb{P}(\{\omega \in \Omega \mid X_n(\omega) \xrightarrow{n \to +\infty} X(\omega)\}) = 1$

3) Convergence in Probability :
$X_n \xrightarrow{\mathbb{P}} X \Leftrightarrow \forall \varepsilon > 0 \ \mathbb{P}(|X_n - X| > \varepsilon) \longrightarrow 0$ as $n \to +\infty$

Related to closeness / consistency of an estimator T :
$\mathbb{P}(|T - \theta| < \varepsilon) \longrightarrow 1 \quad \forall \varepsilon > 0$

4) Convergence in Mean :
$X_n \xrightarrow{r} X \Leftrightarrow \mathbb{E}(|X_n - X|^r) \longrightarrow 0$ as $n \to +\infty$

5) Convergence in Distribution :

$X_n \xrightarrow{d} X \iff \forall x \; F_{X_n}(x) \xrightarrow{n\to+\infty} F_X(x)$ stationarity

6) Slutsky's lemma : If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{\mathbb{P}} a$ then

- $X_n \pm Y_n \xrightarrow{d} X \pm a$
- $X_n \cdot Y_n \xrightarrow{d} X \cdot a$
- $X_n / Y_n \xrightarrow{d} X / a \; , \; a \neq 0$

Special Cases :

- $X_n \xrightarrow{a.s} X \; , \; Y_n \xrightarrow{a.s} Y$ then $X_n + Y_n \xrightarrow{a.s} X + Y$ as $n \to +\infty$
- $X_n \xrightarrow{\mathbb{P}} X \; , \; Y_n \xrightarrow{\mathbb{P}} Y$ then $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$ as $n \to +\infty$

7) Weak Law of Large Numbers : $\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \longrightarrow 0$

8) Strong Law of Large Numbers : $\bar{X}_n \xrightarrow{a.s} \mu$

9) Measurable : $\frac{1}{n} \sum_{i=1}^{n} h(X_i) \xrightarrow{\mathbb{P}} \mathbb{E}(h(X))$

10) Univariate CLT : $\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \sim \mathcal{N}(0,1) \quad S_n = \sum_{i=1}^{n} X_i$

$$\mathbb{P}\left(\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \leq x\right) \longrightarrow \Phi(x)$$

11) Multivariate CLT : $\sqrt{n}\,(\underbrace{S_n - \mathbb{E}(X_1)}_{p\text{-dimension vector}}) \xrightarrow{d} \mathcal{N}(0, \underbrace{\mathrm{Cov}(X_1)}_{p \times p \text{ matrix}})$

12) Continuous Mapping Theorem : $g(X_n) \xrightarrow{\mathbb{P}/as/d} g(X)$ when $X_n \xrightarrow{\mathbb{P}/as/d} X$ and g is continuous